

The Effects of Outcomes-Driven Authentic Assessment on Classroom Quality

Rena Hallam

University of Tennessee

Jennifer Grisham-Brown, Xin Gao, & Robyn Brookshire

University of Kentucky

Abstract

Twenty-six Head Start preschool classrooms participated in a yearlong intervention designed to link the Head Start Child Outcomes Framework with authentic assessment practices. Teachers in intervention and pilot classrooms implemented an assessment approach that incorporated the use of a curriculum-based assessment tool, the development of portfolios aligned with the mandated Head Start Child Outcomes, and the integration of this child assessment information into individual and classroom instructional planning. During the intervention period, comparison classrooms continued to use the assessment approach adopted by the local Head Start program, which included the use of a standardized assessment tool and the use of an agency-developed lesson plan form. Intervention and pilot classrooms demonstrated significant improvements on some dimensions of classroom quality as measured by the Early Language and Literacy Classroom Observation (ELLCO) toolkit, whereas comparison classrooms exhibited no change in classroom quality. Implications for practice are discussed.

Introduction

The application of standards to educational programs as a measure of accountability has become commonplace (National Child Care Information Center, 2006; Scott-Little, Kagan, & Frelow, 2003). In the field of early care and education, this emphasis on standards is often viewed as counter to developmentally appropriate practice and can misguide programs to engage in assessment practices that are not recommended for young children (Meisels, 2000; Neisworth & Bagnato, 2004). This study describes a federally funded project that utilizes the Head Start Child Outcomes Framework as the basis for appropriate authentic assessment practices integrated into instructional planning for young children. This model of outcomes-driven authentic assessment linked to classroom instruction is then examined to determine its effect on classroom quality in preschool programs.

The increasing emphasis on accountability in early care and education programs has illuminated the need for rethinking assessment systems within the field of early childhood education. Increasingly, states and early education entities are developing child standards or child outcomes for children birth through 5 years of age. Head Start specifically developed their Child Outcomes Framework in 2000, outlining the expected outcomes for 4-year-olds as they exit the program (U.S. Department of Health and Human Services, 2003). In response to the Good Start, Grow Smart initiative, states are developing their own set of standards/outcomes for children. The National Child Care Information Center (NCCIC) lists 42 states currently holding or developing early learning guidelines (NCCIC, 2006). A joint position statement of the National Association for the Education of Young Children and the National Association of Early Childhood Specialists in State Departments of Education (NAEYC & NAECS/SDE, 2003) cites both risks and benefits of such early learning standards/outcomes. The potential pitfalls of articulating child standards/outcomes include the negative impact on curriculum and a narrowing of focus on early education activities. However, benefits may also result in that standards may help teachers and programs develop clearer expectations for curriculum and learning goals, facilitate continuity across grade levels, and highlight ways to support children with special needs.

The primary challenge in applying early learning standards or child outcomes to early care and education programs is the potential disconnect between outcomes and appropriate assessment processes for gathering information that can be used at the programmatic level. Head Start, in

particular, has implemented a National Reporting System that uses a standardized assessment process to document child outcomes (Rothman, 2005). Furthermore, the paucity of appropriate assessment tools for young children creates a dilemma in the implementation of a standards-based approach in early care and education. However, the assessment literature in early childhood education underscores the difficulty in obtaining reliable and valid information from young children in standardized assessment formats (Bagnato & Neisworth, 1995; Rafoth, 1997). The appropriate use of assessment tools is also a concern. Many early care and education programs use standardized diagnostic tools for purposes of instructional planning rather than for their intended clinical purpose (Rafoth, 1997). Taken together, these issues emphasize the difficulty in addressing the accountability mandates in the field of early care and education.

Historically, the field of early childhood education has emphasized naturalistic assessment strategies, such as observation and parent interview, as the most appropriate ways to gather meaningful assessment information for young children. Current recommended practices in both early childhood education and early childhood special education focus on authentic assessment approaches. Both the National Association for the Education of Young Children (Bredekamp & Copple, 1997) and the Council for Exceptional Children's Division for Early Childhood (Sandall, McLean, & Smith, 2000) have established guidelines for appropriate assessment practices for young children. These guidelines point to the need for assessment approaches that are developmentally appropriate in terms of the purposes, content, and methods that are used. When assessment is being conducted to support program planning, it should be authentic in that it is ongoing, is conducted in the children's natural contexts, and provides information that is useful in planning for each child.

However, very little empirical research has been conducted on authentic assessment processes. The work of Meisels and colleagues is the exception. Meisels, Liaw, Dorfman, and Nelson (1995) found moderate to high levels of reliability and high predictive validity in the developmental checklist of the Work Sampling System. Additionally, a recent study suggests positive impacts on children's achievement scores in reading and math when teachers use a curriculum-embedded performance assessment system (Meisels, Atkins-Burnett, Xue, Nicholson, Bickel, & Son, 2003).

The intervention approach in this study relied on the use of authentic assessment approaches aligned with the Head Start Child Outcomes Framework. Specifically, Project LINK (*A Partnership to Promote LINKages among Assessment, Curriculum, and Outcomes in Order to Enhance School Success for Children in Head Start Programs*) was a federally funded project that utilized recommended practices in early childhood assessment as a means for documenting accountability. The Assessment, Evaluation, and Planning System (AEPS) for birth to 3 years and 3 to 6 years (Bricker, 2002) was used in the fall and spring to document children's developmental progress. The AEPS is a curriculum-based assessment tool designed to assess children's development and learning across six developmental areas: gross motor, fine motor, social communication, communication, adaptive, and cognitive (Bricker, 2002). In the Project LINK model, the implementation of the AEPS was guided by the use of activity-based protocols. Specifically, six activity-based protocols were developed that, combined with a parent interview and social-communication child observation, complete the full battery of the AEPS. The information gathered from the AEPS was then used to develop individualized child plans for all children enrolled in the Head Start classrooms. After the development of individualized plans, teachers used child assessment data to guide curriculum planning in the classroom (see Figure 1). Additionally, portfolios were developed for all of the children, guided by the mandated Head Start Child Outcomes as well as their individualized goals.

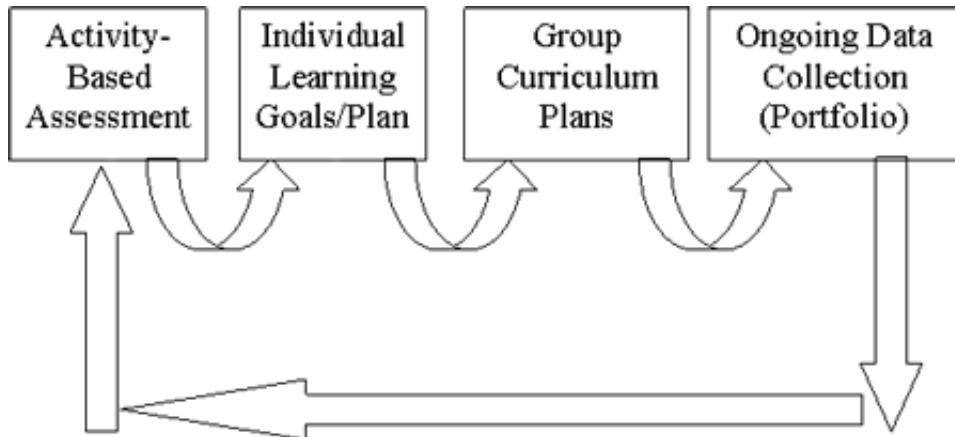


Figure 1. Project LINK Intervention Model.

Comparison classrooms in this study followed current agency procedures for child assessment and curriculum planning. Specifically, classroom teachers used the Learning Accomplishment Profile-Diagnostic (LAP-D, 1992) to collect child assessment data three times a year. Additionally, teachers developed weekly lesson plans using the agency lesson planning format and collected anecdotal observation data on each child in the classroom. Agency policy required teachers to collect one anecdote in each developmental area per child per week.

Conceptually, Project LINK was designed to be a two-year intervention with a two-year evaluation plan; the first year involved examining classroom quality, and the second year involved examining standardized child outcomes and classroom quality. This study outlines the preliminary findings from the pilot and second year of intervention from Project LINK, examining the effects of an outcomes-driven authentic assessment process on classroom quality.

Method

Participants

Project LINK was developed and implemented in partnership with one large multicounty Head Start program consisting of 28 direct-managed preschool classrooms. During the project design, it was determined that first-year teachers would not be included in the project implementation. Thus, only 26 classrooms participated in the project. Given that the classrooms were distributed over multiple sites, classrooms were selected by site to avoid spillover effects. The 26 participating classrooms represented 13 different sites. During the first year of the project, pilot sites were selected in partnership with the administration of the Head Start program. Eight classrooms were selected by Head Start to participate in the pilot portion of the intervention to refine intervention processes and inform model development. The remaining 18 direct-managed preschool classrooms in this Head Start grantee were randomly assigned to intervention and comparison groups by location site and stratified by metropolitan status (urban/rural). The pilot classrooms received the intervention during the pilot year and the subsequent intervention year. During this two-year period, no changes in lead teachers occurred in these classrooms. The intervention group only participated during the targeted second year of intervention. The data presented in Table 1 reflect information collected during the target intervention year.

Table 1
Descriptive Data for Lead Teacher Background Characteristics

Variable	Pilot (n = 8)	Intervention (n = 9)	Comparison (n = 9)
Race			
African American	5 (62.5)	2 (22.2)	2 (22.2)

Caucasian	3 (37.5)	7 (77.8)	7 (77.8)
Education Level			
High School	1 (12.5)		1 (11.1)
AA	2 (25)	3 (33.3)	4 (44.4)
BA	5 (62.5)	5 (55.6)	4 (44.4)
MA		1 (11.1)	
Years of Experience			
1-2 years			
3-5 years	2 (25)	1 (11.1)	3 (33.3)
6-10 years	3 (37.5)	5 (55.6)	3 (33.3)
More than 10 years	3 (37.5)	3 (33.3)	3 (33.3)

Description of the Intervention

Lead teachers, assistant teachers, and children's services coordinators (on-site program managers) attended two days of formal training on the Project LINK model at the beginning of the school year. Training was followed by weekly technical assistance visits throughout the year. The content of the training sessions was designed and delivered by the principal investigators and a specialist in preschool portfolio development. Teachers received instruction and practice on use of the AEPS through the activity-based protocols designed specifically for Project LINK. Teachers were also trained to interpret and utilize AEPS assessment results for developing children's individualized learning plans. Training on the use of a project-specific lesson plan form involved a process of connecting individual assessment results with learning objectives from the Head Start Outcomes Framework. Additionally, teachers were trained to develop a portfolio system for documenting children's ongoing progress toward individualized goals and the Head Start mandated outcomes.

Technical assistance was provided by project staff, all of whom were graduate students in early childhood education. Weekly visits consisted of a variety of supports, including observation and feedback, provision of materials to support implementation of the model, assistance with technology, teacher curriculum resources, and troubleshooting. Visits lasted approximately one hour each week and varied according to the type of assistance provided. Although a range of assistance options were provided to all teachers, the level of help and content of the visits were highly individualized. Teachers with more background in child development and whose prior teaching more closely resembled Project LINK elements may have received more assistance with resources, feedback, and technology, for example; while other teachers with less experience or background knowledge received more direct modeling, observation, and guidance on use of the multiple elements of assessment, lesson planning, and individualization.

All teachers were able to reach at least adequate levels of implementation with the model. AEPS assessments were completed for each child in the fall and spring using activity-based protocols. Individualized learning plans were developed for each child in the classrooms and updated or monitored on a regular basis. Group lesson plans were completed every week, and individual portfolios were created for each child in the classroom to collect evidence of progress throughout the school year.

Measures

Classroom data were collected using the *Early Childhood Environment Rating Scale-Revised Edition* (ECERS-R, Harms, Clifford, & Cryer, 1998) and the *Early Language and Literacy Classroom Observation Toolkit* (ELLCO, Smith & Dickinson, 2002). Inter-rater reliability was

established at 86.72% reliability at the .60 level for the ECERS-R and 100% reliability at a kappa of .60 for the ELLCO. Descriptions of each classroom measure are outlined below.

ECERS-R (Harms, Clifford, & Cryer, 1998) is a widely used program quality measure designed to assess group programs for children of preschool through kindergarten age, 2½ through 5. The scale consists of 43 items organized in 7 subscales (space and furnishings, personal care routines, language-reasoning, activities, interactions, program structure, and parents and staff). The subscale internal consistencies for ECERS-R range from .71 to .88, and the total scale internal consistency is .92 (Harms, Clifford & Cryer, 1998).

ELLCO (Smith & Dickinson, 2002) is a comprehensive set of observation tools designed to describe the extent to which classrooms provide children optimal support for their language and literacy development. The complete toolkit includes three independent research tools (literacy environment checklist, classroom observation, and literacy activities rating scale). The reliability and validity of the three independent tools have been examined. The Cronbach's alpha of .84 for the literacy environment checklist, of .90 for the classroom observation, and of .66 for the literacy activities rating scale show acceptable to good internal consistency (Smith & Dickinson, 2002).

Procedures for Data Collection

Intervention, pilot, and comparison classrooms were observed at the beginning and end of the 2004-2005 year, during scheduled observation times. In the case of the pilot group, this was the second year of implementation of the Project LINK model, given that they had participated on a pilot basis the prior year. The data were collected by seven master's-level and doctoral-level graduate students trained in the implementation of the measures. The ECERS-R data (Harms, Clifford, & Cryer, 1998) and ELLCO data (Smith & Dickinson, 2002) were collected during the same observation period, which lasted from approximately two to four hours. Data collectors scheduled their observations with teachers prior to data collection. Data were entered into SPSS 12.0 for analysis.

Results

Descriptive analyses were conducted for both the ECERS-R (Harms et al., 1998) and the ELLCO (Smith & Dickinson, 2002) for each of the three groups. Table 2 outlines the means and standard deviations for the pretest and posttest scores for the ECERS-R composite and the three ELLCO scales for each of the three groups (intervention, pilot, and comparison). Change scores were then calculated for each of the three groups on all subscale measures (Table 3). ANOVAs were then calculated to examine differences in change scores among the three groups. No statistically significant differences were found relative to the ECERS-R. However, differences in the quality of the language and literacy environment as measured by the ELLCO were found. Specifically, statistically significant differences were found between change scores on the ELLCO Literacy Environment Checklist, $F(2, 23) = 4.82, p < .05$, and the ELLCO Classroom Observation, $F(2, 23) = 10.10, p < .01$. Posthoc analysis using Scheffe indicated that change scores for the pilot group improved more than the comparison group on the ELLCO Literacy Environment Checklist. The pilot group also improved more on the ELLCO Classroom Observation than both intervention and comparison groups. The intervention group improved more than the comparison group on the ELLCO Classroom Observation.

Table 2
Comparison of Pretest and Posttest Means for Classroom Quality

Variable	Pretest Mean <i>M</i> (<i>SD</i>)	Posttest Mean <i>M</i> (<i>SD</i>)
Pilot Group (<i>n</i> = 8)		
ELLCO – Literacy Environment Checklist	22.38 (6.70)	33.50 (7.21)

ELLCO – Classroom Observation	45.00 (5.78)	57.25 (4.89)
ELLCO - Literacy Activity Rating Scale	3.75 (2.61)	6.13 (3.18)
ECERS-R Composite	4.57 (.74)	5.28 (.68)
Intervention Group (n = 9)		
ELLCO – Literacy Environment Checklist	27.56 (10.26)	29.44 (7.50)
ELLCO – Classroom Observation	47.89 (11.68)	56.11 (6.57)
ELLCO - Literacy Activity Rating Scale	6.11 (2.21)	7.78 (2.39)
ECERS-R Composite	4.81 (.43)	5.24 (.47)
Comparison Group (n = 9)		
ELLCO – Literacy Environment Checklist	28.78 (7.81)	26.56 (9.61)
ELLCO – Classroom Observation	50.89 (5.26)	48 (8.70)
ELLCO - Literacy Activity Rating Scale	6.22 (2.05)	5.22 (2.22)
ECERS-R Composite	4.43 (.45)	4.97 (.28)

Table 3
Mean Change Scores for Classroom Quality

Variable	Intervention <i>M</i> (<i>SD</i>)	Pilot <i>M</i> (<i>SD</i>)	Comparison <i>M</i> (<i>SD</i>)	<i>F</i>
ELLCO Literacy Environment Checklist	1.89 (6.37)	11.13 (8.32)	-2.22 (11.49)	4.82*
ELLCO Classroom Observation	8.22 (7.89)	12.25 (6.25)	-2.89 (7.42)	10.10 **
ELLCO Literacy Activity	1.67 (3.35)	2.38 (4.17)	-1.0 (2.18)	2.54
ECERS-R Composite	.44 (.66)	.71 (.85)	.54 (.64)	.32

**p* < .05.

***p* < .01.

Discussion

The infusion of standards into early education programs requires thoughtful planning and reflection on an array of program practices, including child assessment, curriculum planning and implementation, as well as data reporting. The model described in this study is one that allows recommended practices in child assessment to guide these processes while still addressing accountability standards. However, the intent of this study was to examine the impact of an outcomes-driven authentic assessment model on classroom quality and in particular language and literacy environments given the increased focus on language and literacy instruction and the requirements for this emphasis as mandated by the Head Start Child Outcomes Framework.

Findings from this study suggest that an authentic assessment approach may have a positive impact on the language and literacy environment. Differences were found in both the pilot and intervention groups on the ELLCO (Smith & Dickinson, 2002) classroom observation. However, significant differences between pre- and post-observations were not detected on the ECERS-R. It should be noted that all mean ECERS-R scores were below 5.0 (indicative of "good" quality) initially and that the pilot and intervention groups ended the year about 5.0 (5.28 and 5.24, respectively), while the comparison group ended with a 4.97. Given that the ECERS-R is a measure of program quality with an emphasis on structural quality, it is not surprising that differences were not found.

The ECERS-R does not focus on instructional quality and has little emphasis on literacy

instruction in particular (Dickinson, 2002; Stipek & Byler, 2004). Results on the pre/post ELLCO scores suggests that providing a focused intervention on child assessment that is linked to standards and/or a particular content area (such as language and literacy) may result in improved instruction in that area. This improvement in quality related to the use of authentic assessment is consistent with the findings of Meisels et al. (2003), who found that the use of a performance-based curriculum-embedded assessment approach improved child outcomes in primary-age children as measured by standardized achievement test scores. More research is needed in this area to examine how the use of authentic assessment approaches influences teachers' planning of the early childhood curriculum as well as the subsequent impact on child outcomes.

Results from this study should be interpreted cautiously. Data were collected from one large Head Start grantee and therefore cannot be generalized to other types of early care and education programs. Moreover, the pilot and intervention groups received different amounts of exposure to the intervention. Specifically, the pilot programs had participated for two years in the intervention, while the intervention group had only participated for one year. This discrepancy may explain why the data indicated that the pilot group made more gains than the intervention group. The pilot group had more time to digest, learn about, and implement the Project LINK model. Additional study of the impact of authentic assessment on classroom quality needs to be conducted with larger and more diverse samples of programs, teachers, and children. Moreover, because of limited project resources, graduate students were not able to remain blind to treatment groups. Despite these limitations, study results suggest that the use of authentic assessment in early education classrooms may provide an important link to improving classroom quality and curriculum planning.

As the accountability movement unfolds and influences early care and education programs, the potential value of authentic assessment approaches should be systematically examined. Such approaches offer early education programs a means to implement recommended practices in child assessment while continuing to address the growing need to document child outcomes.

References

- Bagnato, Stephan J., & Neisworth, John T. (1995). A national study of the social and treatment "invalidity" of intelligence testing in early intervention. *School Psychology Quarterly, 9*(2), 81-102.
- Bredenkamp, Sue, & Copple, Carol (Eds.). (1997). *Developmentally appropriate practice in early childhood programs* (Rev. ed.). Washington, DC: National Association for the Education of Young Children.
- Bricker, Diane (Ed.). (2002). *AEPS: Assessment, evaluation, and programming system for infants and children: Vol. 2. Test: Birth to three years and three to six years* (2nd ed.). Baltimore, MD: Brookes.
- Dickinson, David K. (2002). Shifting images of developmentally appropriate practice as seen through different lenses. *Educational Researcher, 31*(1), 26-32.
- Harms, Thelma; Clifford, Richard M.; & Cryer, Debby. (1998). *Early childhood environment rating scale* (Rev. ed.). New York: Teachers College Press.
- Learning accomplishment profile—Diagnostic edition.* (1992). Lewisville, NC: Kaplan Early Learning.
- Meisels, Samuel J. (2000). On the side of the child: Personal reflections on testing, teaching, and early childhood education. *Young Children, 55*(6), 16-19.
- Meisels, Samuel J.; Atkins-Burnett, Sally; Xue, Yange; Nicholson, Julie; Bickel, Donna DiPrima; & Son, Seung-Hee. (2003). Creating a system of accountability: The impact of instructional

assessment on elementary children's achievement test scores. Education Policy Analysis Archives, 11(9). Retrieved March 7, 2007, from <http://epaa.asu.edu/epaa/v11n9/>

Meisels, Samuel J.; Liaw, Fong-ruey; Dorfman, Aviva; & Nelson, Regena Fails. (1995). The work sampling system: Reliability and validity of a performance-based assessment for young children. *Early Childhood Research Quarterly, 10*(3), 277-296.

National Association for the Education of Young Children (NAEYC) & National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). (2003). *Early childhood curriculum, assessment, and program evaluation.* Retrieved March 7, 2007, from <http://naeecs.crc.uiuc.edu/position/pisceape.html>

National Child Care Information Center. (2006, May). Selected state early learning guidelines on the Web. Retrieved March 7, 2007, from <http://nccic.org/pubs/goodstart/elgwebsites.pdf>

Neisworth, John T., & Bagnato, Stephan J. (2004). The mismeasure of young children: The authentic assessment alternative. *Infants and Young Children, 17*(3), 198-212.

Rafoth, Mary Ann. (1997). Guidelines for developing screening programs. *Psychology in the Schools, 34*(2), 129-142.

Rothman, Robert. (2005). Testing goes to preschool: Will state and federal testing programs advance the goal of school readiness for all children? *Harvard Education Letter, 21*(2), 1-4.

Sandall, Susan; McLean, Mary; & Smith, Barbara. (2000). *DEC recommended practices in early intervention/early childhood special education.* Longmont, CO: Sopris West.

Scott-Little, Catherine; Kagan, Sharon Lynn; & Frelow, Victoria Stebbins. (2003, June). *Standards for preschool children's learning and development: Who has standards, how were they developed, and how are they used?* Greensboro, NC: SERVE. Retrieved March 7, 2007, from http://www.serve.org/_downloads/publications/Standards2003.pdf

Smith, Miriam W., & Dickinson, David K. (2002). *User's guide to the Early Language and Literacy Classroom Observation toolkit.* Baltimore, MD: Brookes.

Stipek, Deborah, & Byler, Patricia. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly, 19*(3), 375-397.

U.S. Department of Health and Human Services, Administration on Children, Youth, and Families/Head Start Bureau. (2003). *The Head Start path to positive child outcomes.* Washington, DC: Author.

Author Information

Rena A. Hallam, Ph.D., is the executive director of the Early Learning Center for Research and Practice and assistant professor in the Department of Child and Family Studies at the University of Tennessee-Knoxville. She has served in an administrative capacity in both child care and Head Start settings. Her research interests focus on systemic issues related to the quality of early care and education programs, with a particular focus on children living in poverty. Specifically, she has studied transition, assessment and accountability, personnel preparation in early education, and state initiatives to improve child care quality.

Rena Hallam, Ph.D.
Assistant Professor
Interim Executive Director
UT Early Learning Center for Research and Practice
Department of Child and Family Studies
1215 W. Cumberland Avenue, Room 115
Knoxville, TN 37996-1912
Telephone: 865-974-8499
Fax: 865-974-2742
Email: rhallam@utk.edu

Jennifer Grisham-Brown, Ed.D., is an associate professor in the Interdisciplinary Early Childhood Education program at the University of Kentucky (UK), where she teaches undergraduate-, graduate-, and doctoral-level courses. In addition, she is the faculty director of UK's Early Childhood Laboratory, an inclusive birth to 5 program that serves as the primary training site for

UK's early childhood certification program. Dr. Grisham-Brown's research interests include authentic assessment, program quality, and interventions that support the inclusion of young children with disabilities in community-based programs. She has served as an investigator on numerous state and federal grants that have supported her research agenda and has authored or co-authored articles on these topics.

Jennifer Grisham-Brown
University of Kentucky
Department of Special Education and Rehabilitation Counseling
229 Taylor Education Building
Lexington, KY 40506-0001
Telephone: 859-257-8943
Fax: 859-257-1325
Email: jgleat00@uky.edu

Robyn A. Brookshire is a doctoral student in the Department of Child and Family Studies at the University of Tennessee-Knoxville. She served as project coordinator for the Head Start research presented in this article. Her prior experience includes teaching middle school special education and working with at-risk adolescents. She received her master's degree in Interdisciplinary Early Childhood Education at the University of Kentucky.

Robyn A. Brookshire
University of Tennessee
Department of Child and Family Studies
1215 W. Cumberland Avenue
Knoxville, TN 37996-1912
Email: rbrooks8@utk.edu

Xin Gao is research coordinator for the Kids Now Evaluation Project at the University of Kentucky. She brings an extensive background in early childhood education and years of practice in classroom evaluation and child assessment to her current position. She is also a doctoral student in the department of Family Studies, focusing on early childhood education. Her research interests include preschool assessment, children's language and literacy development, and classroom evaluation.

Xin Gao
College of Education, University of Kentucky
Department of Special Education and Rehabilitation Counseling
229 Taylor Education Building
Lexington, KY 40502
Telephone: 859-257-7391
Fax: 859-257-1325
Email: xgao2@uky.edu